*Corresponding Author's Email: Maabdeh@gmail.com
Proceedings of the International Conference on Advanced Research in Management Business and Finance, Vol. 1, Issue. 1, 2024, pp. 29-41
DOI: https://doi.org/10.33422/icmbf.v1i1.897

ISSN: 3030-0649 online

Developing a model to predict defaulting credit card clients using business intelligence techniques

Mohammad Maabdeh*, Ruba Obiedat

The University of Jordan, Jordan

Abstract

The number of personal credit defaults has increased rapidly with the increased popularity of issuing credit cards by banks as a service to customers. Therefore, the risk of defaulters has increased, hence, improving models' performance to predict defaulting credit card clients is receiving more attention recently.

This work aims to develop a model for predicting credit card defaulters by examines the significance of attributes. The results improved by using features selection techniques including change on error approach, filtering infogain attributes, Gain ratio, and Gini algorithms.

In conclusion, this work proposed a model that provides 12 attributes as the most important subset features to assessing the risk of defaulters. The Multilayer perceptron algorithm found to give the best performance among other algorithms tested.

Keywords: Credit card defaulters, Customers behavior, Risk assessment, credit risk, feature selection

1. Introduction

In the modern era of knowledge-based economy, e-commerce, and digital currencies, banking services have grown significantly in both magnitude and function. With this increase in services, the need for risk management and control of service provision has increased as well (Faris, et al., 2020). Therefore, data mining and big data analysis techniques are employed in various ways in the banking firms for various functions, such as identifying potential new credit card customers (Trivedi, 2020) (Nalić, Martinović, & Žagar, 2020) (Devan, Krothapalli, & Shanmugam, 2023), detecting fraud (Desai & Deshmukh, 2013), identifying those potential customers most likely to churn (Muneer, et al., 2022), and many other applications as well (Olson, L, & Delen, 2008).

For the last decade or so, credit cards usage gained increased popularity and the number of users using credit cards has increased significantly, hence defaulting risk has increased as well.

Various methods have been developed to measure and analyze credit risk in financial institutions whose core business is lending (Hamori, et al., 2018). These methods usually require heavy processing and consume high computational power and time. One method to overcome these challenges is to use 'Feature Selection', which attempts to overcome this problem by reducing search space and time. Also, it improves prediction performance, and enables decision makers to better understand existing data sets (Chandrashekar & Sahin, 2013).

This study aims to investigate the significance of data attributes using feature selection methods for predicting credit card defaults, evaluating the impact of different selection techniques on classifier performance. Specifically, we apply Information Gain, Gain Ratio, Gini Index, and Change on Error approaches to determine the most influential attributes in risk prediction. Our experiments compare five machine learning classifiers, Naïve Bayes, Multilayer Perceptron, KNN, Random Forest, and SVM, to identify the best set of attributes that produces a better performance for the prediction system, and provide effective model for credit risk assessment.

The selected attributes for proposed model validated and shows improvement in classification results for most algorithms that used. As a result, this study provides insights into enhancing risk assessment models, allowing institutions to optimize decision making processes and develop more reliable credit approval processes.

The results demonstrate how advanced data driven approaches can mitigate risks and improve credit scoring systems. The remaining of this paper is organized as follows: section 2: Related Work, section 3: Research Methodology, section 4: Data and the techniques used in this study, section 5: Results discussion, and section 6: Conclusions and recommendations.

2. Related Work

Credit-risk management is essential for the bank sector where significant losses can be incurred by financial institutions when borrowers default, because of its importance; a lot of models have been developed aiming to improve defaulters prediction.

Early studies in risk assessment as the other financial models its started on the statistical methods, which is provides the basics of financial and risks assessments (Ozbayoglu, et al., 2020). with improving computational power, the studies shows advancement in presented models with different types of solutions (Moura, et al., 2020), Or combine one or more solutions Such as hyprid or ensemble approches (Alsawalqah, et al., 2017).

With time, many techniques comess to overcome problems , such as **imbalance data**, which is number of defaulters is lower than non-defaulters. (Lei, et al., 2019) provide imbalanced generative adversarial fusion network (IGAFN) model to solve insufuffice data representation, this model take in account both kinds of attributes; static attributes (demograghic information) and dynamic attributes (transaction history).

Many other studies work on **oversambling** (Subasi & Cankurt, 2019) to overcome imbalance data, (Lubis & Huang, 2024) utilize four approches (weighting, SMOTE, Imbalance, and Downsampling) to ensure more comprehensive representation of credit risk.

More advanced techniques comes with increasing complexity of data, **Feature selection** is another critical aspect of credit card default prediction. Researchers have employed various methods to identify the most relevant attributes for improving model performance. (Ibrahim, et al., 2024) evaluated feature selection stratigies, by testing Gini Index, Gain Ratio, and Infrormation Gain, to improve prediction performance.

Building on prior research, this work inhance predecting defulter users by investigating the significant of data attributes, by elemnating irrevelnt data and focousing on impactfull attributes to inhance classification result. Attributes seprated into **static** and **dynamic** attributes, to provide insight into how banks can optimize risk assessment process.

3. Methodology

In this study the problem was formulated as a classification problem to examine whether clients as defaulter or not defaulter users. A comparison was conducted among five classifiers before and after applying feature selection techniques to propose in the final the best set of featuer with best classifier performance. Following are preparing the dataset, feature selection, comparing classifiers, and validation procedures.

3.1 Preparing the Dataset

The default of credit card clients dataset was obtained from the UCI Machine Learning Repository. Its consists of payment and bill records obtained from a bank in Taiwan in the year 2005. it has 30,000 instances, in which 6636 instances for a class of defaultred credit card holders, while the remaning for non defultred users.

To achieve the desired results and to ensure faire comparative among the experiments, the dataset divided into two subsets, 70% for training and 30% for testing. Moreover, a 10-Fold stratified cross validation was applied during training and testing. This approach ensure that a very fair comparative analysis occurred among the experiments, and provide robust validation model prediction.

3.2 Comparison of Classifiers

Various datamining algorithms are to be applied to improve predecting defaulters. five classifers were applied to the original dataset; which is Naïve Bayes, Multilayer Perceptron, KNN, Random Forest, and SVM. The goal was to evaluate prediction performance and to identify the most suitable classifier to measure feature selection effectiveness.

After selecting the classifier with the best performance, the feature selection experiments established to assess attributes importance. Diffirents methods of feature selection were employed based on attributes type. Change on error used for static attributs, while Gini Index, Gain Ratio, and Infrormation Gain used for dynamic attributes.

Once the subset of selected features was determined, the five classifiers were applied to evaluate prediction performance in classifying defaulter with the refined mini dataset. This comparative analysis ensures a systematic approach to improve model performance, and contributing to improve risk assesment.

3.3 Features Selection

Feature selection focouses on most relevent attribute for predicting credit card defults, in aim to reduce the complexity of model, enhance accuracy, and speed up the training process. All that contributes in predictive power of the model.

The features of credit card clients' in the dataset can be divided into two categories (as described in table 3); static features (profile) and dynamic features (behavior), similar approch used by (Lei, et al., 2019), to handle this variation, two methods of feature selections were employed; Change on error, and filtering.

Change on error method was used for the static features. The performance is to be measured before and after dropping a specific attribute, that can identify the contribution of dropped attribute to model performance.

Whereas, **Filtering** method employed to rank importance of dynamic features, using information gain, gan ratio, and gini index algorithms.

3.4 Model Evaluation

Evaluating the performance of a classifier is an important part of model design. It allows for estimating the behavior of the classifier for unseen data, and enabling effective comparison across performance of many classification algorithms.

To measure classification effectiveness in identifying defulters, thre precision (Equation 1) ensures dufulters are truly classified and false positives are reduced. Moreover, recall (Equation 2) is important to detecate all potential defaulters. To balance both factors, the **F-score (Equation 3)** is incorporated, providing an overall evaluation of model effectiveness across both classes.

Table 1: Confusion matrix

		Actual Class	
		Default	Non-Default
Predicted Class	Default	TP (True Positive)	FP (False Positive)
	Non-Default	FN (False Negative)	TN (True Negative)

$$Precision = \frac{Tp}{TP + FP} \tag{1}$$

$$Recall = \frac{Tp}{TP + FN} \tag{2}$$

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

4. Experiments and Results

This section presents the experiments steps, classifier performance evaluation, examine attributes importance, and impact of feature selection techniques on improving credit card default prediction.

4.1 Data Description

The dataset used in this study comprises payment and billing records with a number of 30,000 instances, in which 6636 instances are for a defaulted card holders. The dataset was introduces by (Yeh & Lien, 2009) to predict weather clients would default on their payments, were the datasets represented by two classes “default“ or “not default“.

The dataset includes 25 attributes, which are a mix of continuous and nominal variables, as described in table 2 and table 3. The nominal attributes are encoded by numerical variables, while continuous variables reflect historical payments and billing amounts. The "Pay" attributes capture the past payment behavior, which were traked from April 2005 to September

2005, including the degree of delay in payments for each month (e.g., -1 = payment on time, 1 = payment delayed for one month, 2 = payment delayed for two months, etc.).

Table 2: Attributes Types

Attribute	Type
LIMIT BAL	Numerical
SEX	Categorical
EDUCATION	Categorical
MARRIAGE	Categorical
AGE	Numerical
PAY 1	Categorical
PAY 2	Categorical
PAY 3	Categorical
PAY 4	Categorical
PAY 5	Categorical
PAY 6	Categorical
BILL AMT1	Numerical
BILL AMT2	Numerical
BILL AMT3	Numerical
BILL AMT4	Numerical
BILL AMT5	Numerical
BILL AMT6	Numerical
PAY AMT1	Numerical
PAY AMT2	Numerical
PAY AMT3	Numerical
PAY AMT4	Numerical
PAY AMT5	Numerical
PAY AMT6	Numerical
Default payment next month (The target)	Categorical

Table 3: Attributes description

Attribute	Description
LIMIT BAL	Amount of given credit (NT dollar) for both individual and family.
SEX	Gender (1 = male; 2 = female)
EDUCATION	(1 = graduate school, 2 = university, 3 = high school, 4 = others)
MARRIAGE	Marital status (1 = married; 2 = single; 3 = others)
AGE	In years
PAY 1	Payment status in September
PAY 2	Payment status in August
PAY 3	Payment status in July
PAY 4	Payment status in June
PAY 5	Payment status in May
PAY 6	Payment status in April
BILL AMT1	Amount of bills in September
BILL AMT2	Amount of bills in August
BILL AMT3	Amount of bills in July
BILL AMT4	Amount of bills in June
BILL AMT5	Amount of bills in May
BILL AMT6	Amount of bills in April
PAY AMT1	Amount of previous payment (September)

PAY AMT2	Amount of previous payment (August)
PAY AMT3	Amount of previous payment (July)
PAY AMT4	Amount of previous payment (June)
PAY AMT5	Amount of previous payment (May)
PAY AMT6	Amount of previous payment (April)
Default payment next month	The target (Yes=Default, No=Not default)

The attributes categorized into two types based on the values provided, one as profile attributes and the others as behavior attributes, as detailed in table 4. The profile attributes represent static values such as demographic information, while the behavior attributes represent the dynamic values such as past payments. This approch helps to understand the feature selection experments applied in this study, which was conducted based on the attributes' types.

Table 4: Kinds of attributes

Attribute	Type
LIMIT BAL	Profile
SEX	Profile
EDUCATION	Profile
MARRIAGE	Profile
AGE	Profile
PAY 1-PAY6	Behavior
BILL AMT1-BILL AMT6	Behavior
PAY AMT1-PAY AMT6	Behavior

4.2 Tools Used

Two powerful machine learning tools were utilized to conduct experiments: **WEKA** and **Orange**. These tools provide a range of functionalities for data preprocessing, visualization, and model evaluation.

The University of Waikato, New Zealand, in 1993 founded a tool. Called **WEKA,** that collects a group of machine learning algorithms to solve real world data mining problems which is known as WEKA (Waikato Environment for Knowledge Analysis), written in Java, and runs almost on any platform. It is free software and available on WEKA website. The WEKA workbench contains a collection of visualization tools besides algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to the functions (Ibrahim & Shiba, 2019). WEKA allows users to quick access to different machine learning methods. The tool includes algorithms to solve regression, classification, clustering, association rule mining and attribute selection problems (Hall, et al., 2009).

Orange is a machine learning and data mining software that uses data analysis based on Python scripting and visual programming. Orange is provided for experienced users, programmers, and students with a graphical user interface containing a set of packages for data visualization, machine learning, data mining, and data analysis. Orange tool is easier to use and has a friendly user interface with drag and drop options (Demˇsar, Curk, & Erjavec, 2013).

4.3 Experiments

4.3.1 Comparison of Classifiers' Performances

For each classifier, there is a special way of problem solving. In the first step, five classifers were applied to the original dataset;

Naïve Bayes:

Multilayer Perceptron

KNN

Random Forest

SVM

To ensure a fair comparison, 10-fold stratified cross-validation was applied. Table 5 summarize the weighted F-measure for each classifier. Additionally, table 6 reports accuracy metrics by class:

Table 5: Compare among Classifiers-Weighted average

	Naïve Bayes	M.L Perceptron	KNN	Random Forest	SVM
WT. F-Measure	0.69	0.818	0.715	0.819	0.470

Table 6: Compare among Classifiers - Accuracy measured By Class

		Naïve Bayes	M.L Perceptron	KNN	Random Forest	SVM
F-Measure	Default	0.657	0.347	0.240	0.380	0.323
	Non-Default	0.704	0.952	0.850	0.944	0.512
Recall	Default	0.487	0.457	0.179	0.482	0.612
	Non-Default	0.782	0.891	0.912	0.891	0.382
Precision	Default	0.310	0.389	0.366	0.403	0.219
	Non-Default	0.310	0.389	0.796	0.403	0.776

Among the classifiers, Random Forest demonstrated superior predictive capabilities, followed by Multilayer Perceptron. The descending order of classifier performance is as follows

1. Random Forest
2. Multilayer Perceptron
3. Naïve Bayes
4. KNN
5. SVM

To further refine performance, feature selection techniques were applied, and classifier comparisons were reevaluated using selected attributes.

4.3.2 Feature selection

Feature selection was employed to minimize redundant attributes while retaining the most influential attributes. The following techniques were utilized:

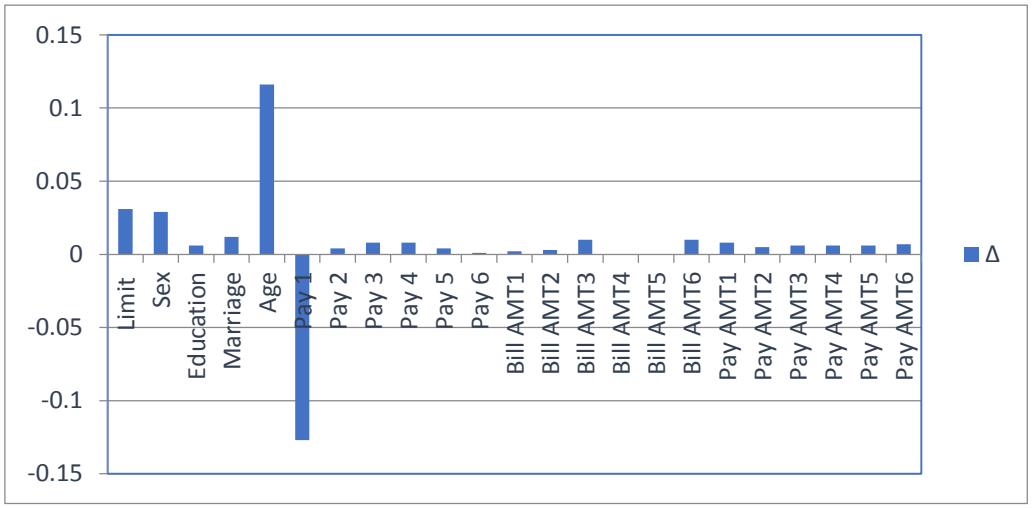
Change on Error

This technique assesses how attribute removal affects classification precision, this change (Δ) could be positive or negative. A neural network model with 10-fold cross-validation was used to measure precision changes (Δ) after remove each attribute separately. The results detailed in table 7.

Table 7: Change on error

Att.	Sex	Education	Marriage	Age	Pay 1	Pay 2	Pay 3
Δ	0.029	0.006	0.012	0.116	-0.127	0.004	0.008
Att.	Pay 4	Pay 5	Pay 6	Bill AMT1	Bill AMT2	Bill AMT3	Bill AMT4
Δ	0.008	0.004	0.001	0.002	0.003	0.01	0
Att.	Bill AMT5	Bill AMT6	Pay AMT1	Pay AMT2	Pay AMT3	Pay AMT4	Pay AMT5
Δ	0	0.01	0.008	0.005	0.006	0.006	0.006

Figure 1: Change on error



As explained in figure 1, removing Pay 1 decrease the precision value, which indicates the significant value of this attribute in model design. In contrast, removing "**Age**" had minimal effect.

Filter

A filtering methos was applied to rank attributes by importance. By using Information Gain, Gain Ratio, and Gini Index. The ranked attributes are listed in Figure 2.

Figure 2: Attributes ranking

Attributes	Info. Gain	Gain ratio	Gini
PAY_1	0.109720759	0.052937700	0.061626519
PAY_2	0.070772982	0.038402375	0.039903223
PAY_3	0.053780934	0.029459342	0.030118201
PAY_4	0.047447060	0.026760156	0.026891089
PAY_5	0.044065100	0.025486803	0.025239875
PAY_6	0.038119240	0.021587254	0.021669745
LIMIT_BAL	0.019700976	0.009854519	0.009406159
PAY_AMT1	0.017305553	0.008652808	0.008256609
PAY_AMT2	0.015725644	0.007864507	0.007394718
PAY_AMT3	0.013880520	0.006940270	0.006586882
PAY_AMT4	0.011398862	0.005699777	0.005395421
PAY_AMT6	0.010631284	0.005316365	0.004958750
PAY_AMT5	0.009306691	0.004653369	0.004369470
EDUCATION	0.004441289	0.002798633	0.001874494
AGE	0.003702893	0.000727959	0.001820933
BILL_AMT6	0.001443966	0.000721983	0.000697363
BILL_AMT4	0.001209354	0.000604677	0.000580980
SEX	0.001144122	0.001181060	0.000550180
MARRIAGE	0.000880265	0.000809120	0.000409572
BILL_AMT1	0.000628177	0.000314088	0.000295591
BILL_AMT5	0.000609683	0.000304841	0.000292110
BILL_AMT3	0.000461141	0.000230570	0.000218756
BILL_AMT2	0.000449260	0.000224630	0.000211804

The numbers illustrate the significance of "Pay attributes", with "Pay 1" consistently ranking first across all filters.

Applying Feature Selection

On can see that, past payment behavior plays a crucial role in predicting credit card defaults. The dataset includes six months' history of transaction-related attributes, for Pay, Bill-amt, and bill_amt records. However, data visualized in figure 2 suggest that more recent financial behavior tends to have a stronger correlation with default risk.

To enhance model efficiency, we retained only the most recent three months of dynamic attributes, discarding older transaction records. This approach ensures that classifiers focus on fresh behavioral patterns, reducing noise and improving predictive accuracy.

After applying feature selection techniques, only the top ranked attributes were retained for classification. The classifiers were reevaluated using the refined feature subset.

Classifier Comparison After Feature Selection

After selecting the most significant attributes, the five classifiers were retrained and tested, comparison results summarized in table 8:

Table 8: Comparison among Classifiers for subset attributes

	Naïve Bayes	M.L Perceptron	KNN	Random Forest	SVM
WT. F-Measure	0.766	0.819	0.714	0.812	0.304

After feature selection process, Multilayer Perceptron leds best performing model, surpassing Random Forest, which initially led the comparison.

Effect of Feature Selection on Classification Performance

The difference can be noticed before and after choosing a subset of features, table 9 summariz results.

Table 9: Performance - before and after choosing subset of features

Attributes	Class	Precision	Recall	F-Measure	WT. F-Measure
Original features	Default	0.389	0.457	0.347	0.818
	Non-Default	0.389	0.891	0.952	
Subset Features	Default	0.393	0.458	0.345	0.819
	Non-Default	0.393	0.891	0.954	

The results indicate that removing less important attributes did not negatively impact model performance, demonstrating the efficiency of feature selection in reducing complexity while maintaining predictive performance.

5. Discussion

In aims of increasing the performance of risk assessment for card holders, first step of comparing shows that the comparative analysis of classifiers demonstrated that Random Forest initially performed best when using the full dataset. After that two techniques of feature selection were applied; change on error and filtering. The attributes are classified into profile attributes and behavior attributes, where selecting any feature of behavior attributes necessitates choosing the alternatives of other dynamic attributes for the same months. However, after applying feature selection techniques, the Multilayer Perceptron surpassed Random Forest in predictive accuracy.

The best set of attributes, as shown in table 10, shows that the historical payments attributes are the most influential in predicting credit card default risk. Feature selection confirms that retaining only the most recent three months of transaction data enhances model reliability

After removing the neglected attributes, evaluating the performance gives approximately the same performance for the original set of features. This indicates that we can trust these features to assess the risk.

Table 10: subset features

No.	Attribute
1	LIMIT BAL
2	PAY ١
3	PAY 2
4	PAY 3
5	BILL AMT1
6	BILL AMT2
7	BILL AMT3
8	PAY AMT1
9	PAY AMT2
10	PAY AMT3

These findings add valuable implications for financial institutions. Banks can enhance their credit risk management systems by prioritizing recent transaction history and key behavioral attributes when assessing applicants. Eliminating redundant attributes reduces computational load, and ensures efficient data processing for approval processes.

6. Conclusion

Popularity of using credit card boosted the need for risks assessment of customers to default. This paper presented a model to enhance the prediction accuracy for classification of credit card customers as defaulters or non-defaulters using a dataset from the UCI Machine Learning repository.

This model developed by identifying the most significant attributes, by using change on error and Filtering methods, after separating features for two kinds, first as static (profile) features and the second as dynamic (behavior) features.

As a result of the conducted experiments, it has been found that the recommended algorithm for default credit classification is multilayer perception after find the best 12 subset of features.

After comparing the results of both the original dataset and the subset of features, the results were approximately the same with a little improvement for the subset side. Nevertheless, the results of the subset dataset showed significant saving in terms of computational time, reduced complexity, and reduced overfitting.

This model provides a valuable insight that enables banks to review their risk management strategies, moreover helps in credit approval decisions.

References

Alsawalqah, H., Faris, H., Aljarah, I., Alnemer, L., & Alhindawi, N. (2017). Hybrid SMOTE-ensemble approach for software defect prediction. In Software Engineering Trends and Techniques in Intelligent Systems: Proceedings of the 6th Computer Science On-line Conference 2017 (pp. 355-366). Springer. https://doi.org/10.1007/978-3-319-57141-6_39

Chandrashekar, G., & Sahin, F. (2013). A survey on feature selection methods. Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, 40, 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Demˇsar, J., Curk, T., & Erjavec, A. (2013). Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research, 14, 2349-2353.

Desai, A. B., & Deshmukh, R. (2013). Data mining techniques for Fraud Detection. International Journal of Computer Science and Information Technologies, 4, 1-4.

Devan, M., Krothapalli, B., & Shanmugam, L. (2023). Advanced Machine Learning Algorithms for Real-Time Fraud Detection in Investment Banking: A Comprehensive Framework. Cybersecurity and Network Defense Research (CNDR), 3, 57-94.

Faris, H., Abukhurma, R., Almanaseer, W., Saadeh, M., Mora, A. M., Castillo, P. A., & Aljarah, I. (2020). Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market. Progress in Artificial Intelligence, 9, 31-53. https://doi.org/10.1007/s13748-019-00197-9

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18. https://doi.org/10.1145/1656274.1656278

Hamori, S., Kawai, M., Kume, T., Murakami, Y., & Watanabe, C. (2018). Ensemble Learning or Deep Learning? Application to Default Risk Analysis. Risk and Financial Management. https://doi.org/10.3390/jrfm11010012

Ibrahim, F. A., & Shiba, O. A. (2019). Data Mining: WEKA Software (an Overview). Journal of Pure & Applied Sciences, 18, 54-58.

Ibrahim, N., Ishak, U. M., Ali, N. N., & Shaadan, N. (2024). Machine learning-based approaches for credit card debt prediction. Malaysian Journal of Computing (MJoC), 9(1), 1722-1733. https://doi.org/10.24191/mjoc.v9i1.25656

Lei, K., Xie, Y., Zhong, S., Dai, J., Yang, M., & Shen, Y. (2019). Generative adversarial fusion network for class imbalance credit scoring. Neural Computing and Applications. https://doi.org/10.1007/s00521-019-04335-1

Lubis, R. M., & Huang, J.-P. (2024). Leveraging Machine Learning to Predict Credit Card Customer Segmentation. Journal of Ecohumanism, 3(7), 3386-3418.

Moura, J., Bissaro, L., Santos, F., & Carneiro, M. (2020). Deep Learning in Risk Assessment. Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, 1068-1079. https://doi.org/10.5753/eniac.2019.9358

Muneer, A., Ali, R. F., Alghamdi, A., Taib, S. M., Almaghthawi, A., & Ghaleb, E. A. (2022). Predicting customers churning in banking industry: A machine learning approach. Indonesian Journal of Electrical Engineering and Computer Science, 26, 539. https://doi.org/10.11591/ijeecs.v26.i1.pp539-549

Nalić, J., Martinović, G., & Žagar, D. (2020). New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. Advanced Engineering Informatics, 45. https://doi.org/10.1016/j.aei.2020.101130

Olson, L, D., & Delen, D. (2008). Advanced data mining techniques. Springer Science & Business Media.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. Applied soft computing, 93. https://doi.org/10.1016/j.asoc.2020.106384

Subasi, A. a., & Cankurt, S. (2019). Prediction of default payment of credit card clients using Data Mining Techniques. Fifth International Engineering Conference on Developments in Civil & Computer Engineering Applications, (pp. 115-120). Erbil-IRAQ. https://doi.org/10.1109/IEC47844.2019.8950597

Trivedi, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. Technology in Society, 63. https://doi.org/10.1016/j.techsoc.2020.101413

Yeh, C., & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, 36, 2473-2480. https://doi.org/10.1016/j.eswa.2007.12.020